

Anwendungsfall Narragonien

Sebastian Brants „Narrenschiff“ gilt als Schlüsseltext der Frühen Neuzeit. Die Erstausgabe von 1494 präsentiert eine bunte Revue von 109 Narrentypen, die je für ein Laster stehen und in Holzschnitt und Spruchgedicht vor Augen gestellt werden.

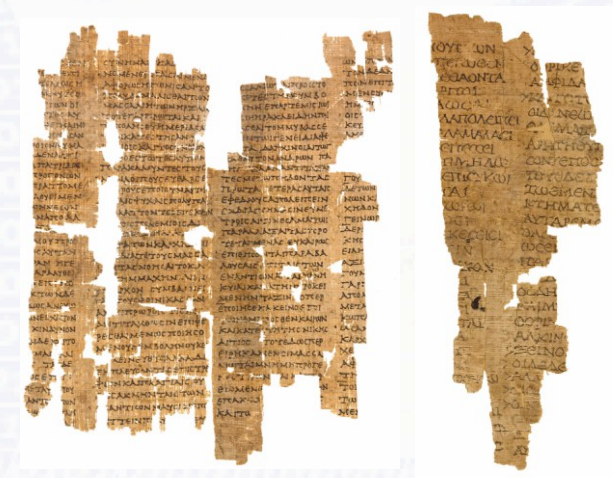


Das „Narrenschiff“ zählt zu den ambitioniertesten intermedialen Experimenten des Frühdrucks. Den enormen Erfolg auf dem europäischen Buchmarkt belegen zahlreiche Neuauflagen, Nachdrucke und Übertragungen ins Lateinische und in die Volkssprachen. „Narragonien digital“ will die Textualität, Medialität und Überlieferungsgeschichte des „Narrenschiffs“ in einer digitalen Edition von 8 Fassungen vor 1500 abbilden. Sie soll das komplexe Seitenlayout veranschaulichen, die historischen Text-, Bild- und Layout-Transformationen dokumentieren und die geschichtliche Eigenbewegung des Werkes erläutern. Spezifische OCR-Verfahren und synoptische Editionstools werden dazu entwickelt.

Leitung: Prof. B. Burrichter, Prof. J. Hamm
Mitarbeiterinnen: Christine Grundig, Viktoria Walter

Anwendungsfall Anagnosis

Ziel von Anagnosis ist eine automatisierte, buchstabengenaue Verknüpfung zwischen den im Digital Corpus of Literary Papyri gesammelten Texten und den dazugehörigen Abbildungen. Die Software kombiniert die Ergebnisse einer Segmentierung der Bilder und die vom Transkript bereitgestellten Informationen: Alle Buchstaben auf einem Scan werden durch Boxen markiert und den entsprechenden Zeichen der Transkription zugeordnet.



Die Glyphenboxen werden durch Koordinaten identifiziert: Dies ermöglicht das benutzerdefinierte Hochladen zusätzlicher Abbildungen (z. B. von Multispektralaufnahmen desselben Papyrus) ohne neue Segmentierung. Schließlich soll die Software die Erzeugung wiederkombinierbarer Zeichensets gewährleisten und somit ein nützliches paläographisches Werkzeug zur Verfügung stellen.

Leitung: Dr. Holger Essler
Mitarbeiter: Vincenzo Damiani

Anwendungsfall Schulwandbilder digital

Das Teilprojekt erschließt und digitalisiert den Bestand schulischer Wandbilder sowie didaktischer Handreichungen und Lehrerzeitschriften der Forschungsstelle Historische Bildmedien Würzburg. Diese Materialien prägten die Welt- und Wirklichkeitsvorstellung der Schülergenerationen zwischen 1840-1990.



Aus heutiger Sicht sind sie Zeugnisse historisch-pädagogischer Denkfiguren sowie kultureller und politischer Einflüsse und liefern damit einen bedeutenden Beitrag zur interdisziplinären Forschung. Das Projekt erprobt ein Erschließungskonzept für Bildquellen und nicht primär textorientierte Universitätssammlungen. Die für die Begleitmaterialien zu entwickelnden Textanalysetools dienen im Forschungsverlauf als Grundlage für komplexere semantische Analysen und zur Gewinnung von Deskriptoren, die im Sinne einer „Verschlagwortung“ Schulwandbilder thematisch recherchierbar machen.

Leitung: Prof. Dr. A. Dörpinghaus, Dr. Ina Uphoff
Mitarbeiter: Christian Rappold

KONTAKT Kallimachos

DH-Zentrum - Digitalisierungszentrum
Universitätsbibliothek Würzburg
Am Hubland, D-97074 Würzburg
<http://www.kallimachos.de>
Tel. +49 (0)931 31 80534
kallimachos@bibliothek.uni-wuerzburg.de

Projektleitung
Dr. Hans-Günter Schmidt, Universitätsbibliothek
hans-guenter.schmidt@bibliothek.uni-wuerzburg.de

Quantitative Verfahren zur Datenanalyse, Textmining
Prof. Dr. Fotis Jannidis, Lehrstuhl für Neuere Deutsche Literatur und Computerphilologie
fotis.jannidis@uni-wuerzburg.de
Prof. Dr. Frank Puppe, Lehrstuhl Informatik VI
frank.puppe@uni-wuerzburg.de

OCR
Prof. Dr. Andreas Dengel, Deutsches Forschungszentrum für Künstliche Intelligenz, Kaiserslautern
andreas.dengel@dfki.de

Computerlinguistik
Prof. Dr. Stefan Evert, Universität Erlangen
stefan.evert@fau.de

Alle weiteren Projektbeteiligten finden Sie bei den einzelnen Teilprojekten.

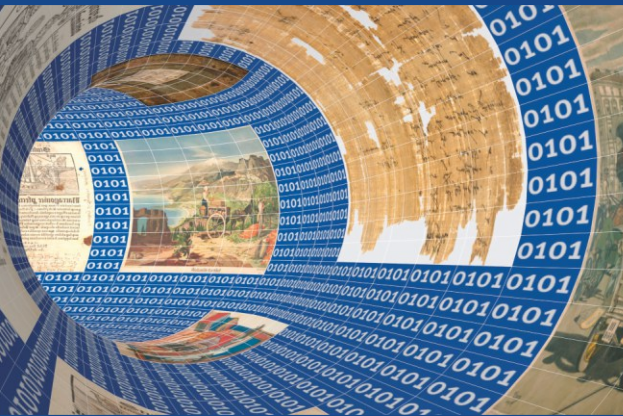


(Stand: Januar 2016)

Universitätsbibliothek

Kallimachos

Zentrum für digitale Edition und quantitative Analyse



GEFÖRDERT VOM



Kallimachos

<http://www.kallimachos.de>

KALLIMACHOS

Kallimachos baut an der Universitätsbibliothek Würzburg ein Digital Humanities Zentrum mit den Arbeitsschwerpunkten OCR, digitale Edition und quantitative Analyse auf. Das Zentrum bietet eine technische und soziale Infrastruktur, die Geisteswissenschaftler bei der Beantwortung von Forschungsfragen unterstützt und innovative Werkzeuge, Methoden und Dienste entwickelt:

- Open-Source-Komponenten („Module“), die in bereits vorhandene Systeme integriert werden.
- Prototypische Arbeitsabläufe, in deren Fokus die Anwendung quantitativer Verfahren in der kompletten Prozesskette textwissenschaftlicher Forschung steht
- Verfahren zur Online-Präsentation, Archivierung und Vernetzung von Textkorpora und digitalen Sammlungen.
- Angebote zum Kompetenzaufbau: Online-Tutorials, Schulungen, Workshops.

In 6 Anwendungsfällen aus Papyrologie, Geschichte, Philologie, Philosophie und Pädagogik werden die Neuentwicklungen erprobt und die Fruchtbarkeit der neuen Verfahren demonstriert. Zugleich werden Bausteine für elektronisches Forschungsdatenmanagement in den Geisteswissenschaften entwickelt.



OCR-Modul

Kallimachos geht neue Wege der optischen Schrifterkennung. An der Universitätsbibliothek Würzburg wird im Rahmen eines klassischen, auf der OCR-Engine Tesseract basierenden Workflows die Verbesserung von Erkennungsergebnissen durch detaillierte Erfassung von Drucktypen und verlagsspezifisches Mustertraining erprobt.

Dyß puchlein wurt güt zu dem Kauff

Dyß puchlein wurt güt zu dem kauff

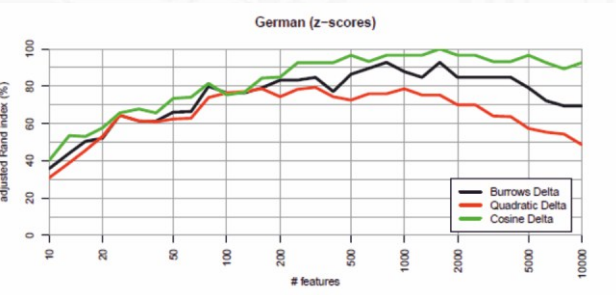
Die Forschung am DFKI Kaiserslautern geht einen Schritt weiter: Unter Fortentwicklung der Software OCROPUS und von Konzepten künstlicher neuronaler Netze soll „anyOCR“ sich automatisch an die Probleme historischer Druckschriften adaptieren – ohne manuell erstellte Gutmuster, Typeninventare und korrigierendes intellektuelles Eingreifen durch Philologen. Hinzu tritt die Weiterentwicklung von Methoden für Bildvorverarbeitung, Analyse von Layout und Sprachmodellierung.

Editions- und Wiki-Modul

Neben etablierte XML-Editoren mit ihrer Funktionsvielfalt für den Spezialanwender treten webbasierte schlankere, intuitiver bedienbare Werkzeuge, die über eine Nutzer-, Rollen- und Rechteverwaltung fest in eine Redaktionsumgebung eingebettet sind. Kollaborative OCR-Korrektur, Editionserstellung und -pflege soll damit sowohl intern, für Arbeitsgruppen an Lehrstühlen, als auch extern über Wiki-Systeme erleichtert werden.

Modul Datenanalyse

Eine Edition ist für Kallimachos nicht Endstation, sondern Zwischenstopp zu weiteren Analysen. Der Fokus liegt dabei auf der Entwicklung und Anpassung regelbasierter Verfahren und Verfahren des maschinellen Lernens für die Arbeit mit literarischen Korpora. Dazu entwickelt das Modul anhand der Anwendungsfälle „Leserlenkung“, „Narrative Techniken“ und „Identifizierung anonymer Übersetzer“ Techniken zur Zusammenstellung eines Untersuchungskorpus, zu dessen Aufbereitung für Textanalyseverfahren, für die Analyse selbst und für die Einbindung der Analyseergebnisse in TextGrid und andere Repositorien.

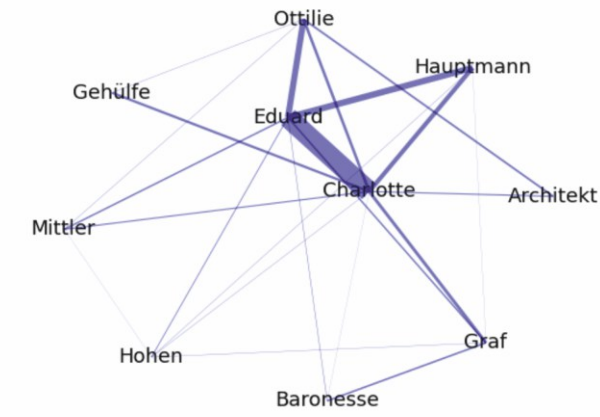
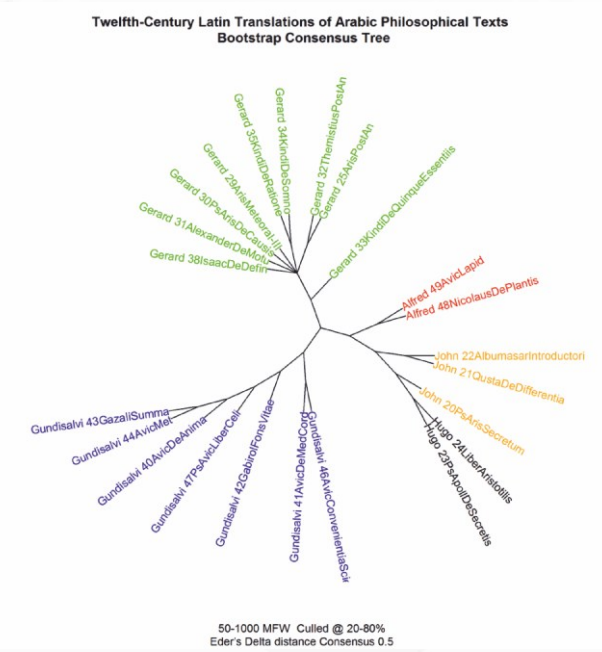


Modul Prototypische Abläufe

Die neu entwickelten Arbeitsabläufe werden formalisiert, verständlich dokumentiert und in Form von Tutorials zur Nachnutzung mit eigenen Datenbeständen zur Verfügung gestellt. Dabei werden sowohl etablierte Tools, als auch Eigenentwicklungen eingebracht, Erfahrungen und Ergebnisse präsentiert und damit Best Practices zur computergestützten Analyse von geisteswissenschaftlichen Daten etabliert.

Anwendungsfälle Leserlenkung - Narrative Techniken

An Textkorpora aus Germanistik und Romanistik wird eine Vielzahl von quantitativen Analysemethoden erprobt. Gegenstand der Untersuchungen sind u.a. Romanfiguren und ihre Charakterisierung, Figurenkonstellationen und -Beziehungen. Daneben findet Forschung zur automatischen Klassifikation von literarischen Gattungen, Autorschaftsattribuion, sowie zu Weiterentwicklung und besserem Verständnis stilometrischer Methoden statt. Schließlich sollen in diesem Kontext valide Methoden zur Signifikanzüberprüfung quantitativer Daten entwickelt werden, die den Besonderheiten literarischer Korpora (kleine Stichprobenumfänge, hohe sprachliche Variabilität, fehlerbehaftete automatische Analysen) Rechnung tragen.



Anwendungsfall Identifizierung anonymer Übersetzer

Die Übersetzung zahlreicher Texte im 12. Jh. aus dem Arabischen ins Lateinische beeinflusste die Entwicklung der europäischen Wissenschaften aller Fachgebiete maßgeblich. Dennoch sind viele der Übersetzungen nur anonym überliefert.



Im Zuge des Projektes sollen philosophische, astronomische und astrologische Texte zunächst digitalisiert werden, um anhand dieses Textcorpus charakteristische stilistische Merkmale zur Identifizierung anonymer Übersetzer mit computer-gestützten stilometrischen und philologischen Methoden ermitteln zu können.

Leitung: Prof. Dr. Dag Nikolaus Hasse
Mitarbeiter: Andreas Büttner